

Introduction to **Linked Open Data**

Mustafa Jarrar

Birzeit University



Watch this lecture and download the slides



Online Courses : <http://www.jarrar.info/courses/>

Outline

- What/why Linked Open Data
- Web of Data
- Interlinked structured data.
- Classical Web
- The challenge
- Web APIs and Mashups
- How to create linked-data?
- Properties of the Web of Linked Data
- Linked Data

Keywords: Semantic Web, Data Web, Web 3.0, Classical Web, Web 3.0, Web APIs , Mashups, Linked Data

Linked Open Data - What is it?



A nice video by EuropeanaEU in YouTube
<http://www.youtube.com/watch?v=uju4wT9uBIA>

Semantic/ Data Web/ Web 3.0?

“The goal of the Semantic Web is to create a universal medium for the exchange of **DATA**”, W3C.



“The Semantic Web is a **web of data**, in some ways like a **global database**”,
Tim Berners-Lee – Inventor of the WWW.



Web of Data

The Data Web envisions the web as a world-wide interlinked structured data.

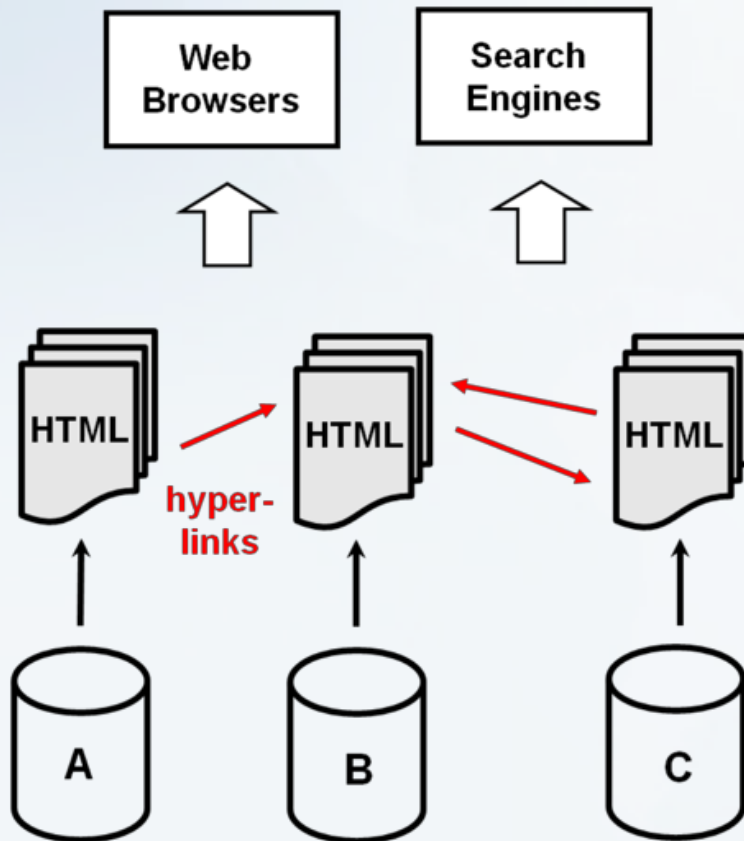
The Web as we know it today is a global information space of linked documents.

The same vision is applied to data: publishing and connecting structured data on the web.



Classical Web

Based on [1]



The classical web is a global information space of linked documents.

Primary Units of the hypertext Web are:

- HTML Documents,
- Connected by Hyperlinks

The challenge

Based on [1]

The problem is that the information on the classical web is not structured.

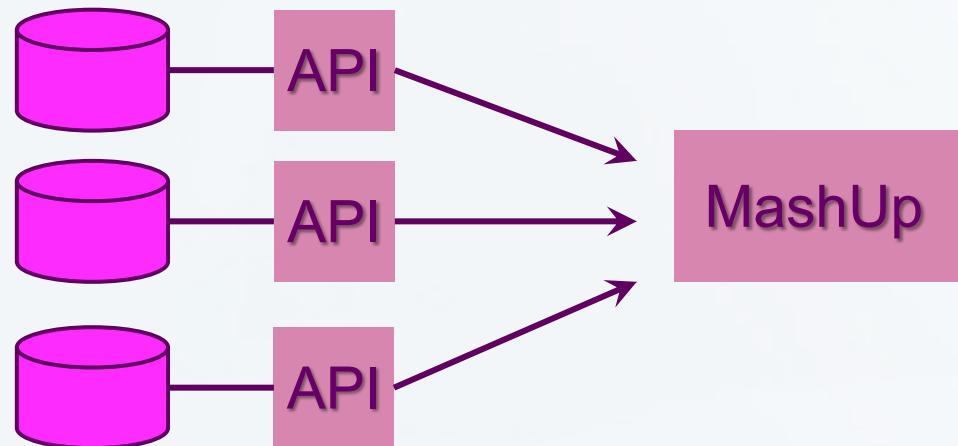
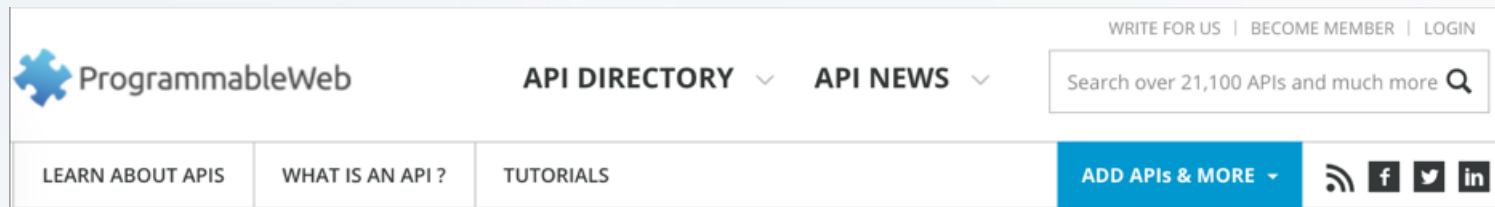
- Programs cannot use such information in a useful way.

The solution is to increase the structure of published information.

Web APIs and Mashups

Many major data sources such as (Amazon, Yahoo!, eBay, Google, WikiData, OpenStreetMap, GeoNames,...) provide access to their data through APIs.

Currently, programmableweb.com lists 21,100 APIs (up to 14. March 2019).



Web APIs and Mashups

The challenges are:

- APIs provide proprietary interfaces,
- Data retrieved from these APIs is represented using different formats (different data models).
- Mashups created using these APIs are based on a fixed set of data sources. This is because entities in different APIs are not linked.
- You cannot set hyperlinks between entities.



From [1]

Jarrar © 2018

APIs separates data

Beyond Web APIs and Mashups: The Data Web and Linked Data

The Data Web envisions the web as a world-wide interlinked structured data.

Linked data refers to the set of best practices for publishing and connecting structured data on the web.

Linked data best practices has lead to the extension of the web connecting data from diverse domains such as:

People, companies, land marks, locations, events, books, scientific publications, films, music, television and radio programs, genes, proteins drugs, clinical trials, online communities, statistical and scientific data, reviews, ...

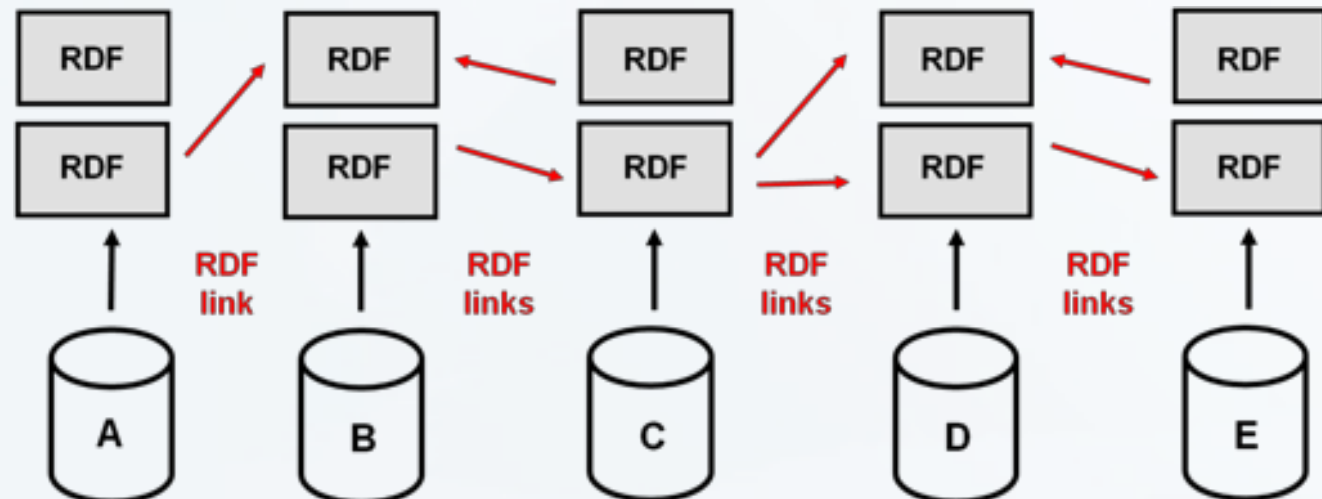
The Data Web and Linked Data

Based on [1]

While the primary units of the hypertext Web are HTML documents connected by un-typed Hyperlinks, Linked Data relies on documents containing data in RDF.

However, rather than simply connecting these documents, Linked Data uses RDF to make typed statements that link arbitrary things in the world.

The result is a web of things in the world, described by data on the Web



How to create linked-data?

The four principles (By Tim Berners-Lee)

Berners-Lee (2006) outlined a set of 'rules' for publishing data on the Web in a way that all published data becomes part of a single global data space:

1. Use URIs to denote things. Most things are referenced by unique URIs.
2. Use HTTP URIs so that these things can be referred to and looked up ("dereferenced") by people and user agents.
3. Provide useful information about the thing when its URI is dereferenced, leveraging standards such as RDF, SPARQL.
4. Include links to other related things (using their URIs) when publishing data on the Web.

Properties of the Web of Linked Data

Based on [2]

Anyone can publish data to the Web of Linked Data

Entities are connected by links

- Creating a global data graph that spans data sources and enables the discovery of new data sources.

Data is self-describing

- If an application encounters data represented using an unfamiliar vocabulary, the application can resolve the URIs that identify vocabulary terms in order to find their RDFS or OWL definition.

The Web of Data is open

- Meaning that applications can discover new data sources at run-time by following links.

The Linked Open Data Cloud

A community project to publish various open datasets as RDF on the Web and by setting RDF links between data items from different data sources.

Data items are linked across datasets (Palestine in DBPedia is the same Palestine in Yago, in Fliker, YouTube,)



Linking Data

How are same entities described in different datasets linked?

By linking the Global Identifier, that is, the URI!

Let's have a look at real examples from real datasets:

```
<http://dbpedia.org/resource/Bethlehem> owl:sameAs  
<http://sws.geonames.org/284315/>
```

- Linking the entity “Bethlehem” between the DBPedia dataset and the Geonames dataset in the Linking Open Data cloud.
- This is done by linking the URIs of “Bethlehem” in both datasets using **owl:sameAs**.

```
<http://dbpedia.org/resource/Tim_Berners-Lee> owl:sameAs  
<http://www4.wiwiss.fu-berlin.de/dblp/resource/person/100007>
```

- Linking the entity “Tim Berners-Lee” between the DBPedia dataset and the DBLP dataset .
- This is done by linking the URIs of “Tim Berners-Lee” in both datasets using **owl:sameAs**.

Resources

About: Bethlehem

An Entity of Type: [populated place](#), from Named Graph: <http://dbpedia.org>, within Data Space: [dbpedia.org](#)

Bethlehem is a Palestinian city in the central West Bank, approximately 5 kilometers (5.0 mi) south of Jerusalem, with a population of about 30,000 people. It is the capital of the Bethlehem culture and tourism. The Hebrew Bible identifies Bet Lehem as the city David was from and the location where he was crowned as the king of Israel.

Property	Value
dbpedia:enl:abstract	<ul style="list-style-type: none">Bethlehem ist eine Stadt im Westjordanland mit 29 530 Einwohnern. Die Stadt gehört zu den palästinensischen Autonomiegebieten und grenzt im Agglomeration Bethlehem gehören auch Beit Dschala und Beit Sahur, letzterer Ort hat wie Bethlehem biblische Bedeutung. Bürgermeister der Stadt pensionierte Arzt Victor Batarseh. Für die 2.1 Mrd. Christen ist die Stadt von besonderer Bedeutung, weil sie der Überlieferung nach als der Gebort Bethlehem is a Palestinian city in the central West Bank, approximately 5 kilometers (5.0 mi) south of Jerusalem, with a population of about 30,000 National Authority and a hub of Palestinian culture and tourism. The Hebrew Bible identifies Bet Lehem as the city David was from and the location of Matthew and Luka identify Bethlehem as the birthplace of Jesus of Nazareth. The town is inhabited by one of the oldest Christian communities: emigration. The city was sacked by the Samaritans in 529 AD, during their revolt, but was rebuilt by the Byzantine emperor Justinian I. Bethlehem guaranteed safety for the city's religious shrines. In 1099, Crusaders captured and fortified Bethlehem and replaced its Greek Orthodox clergy with Saladin, the sultan of Egypt and Syria. With the coming of the Mamluks in 1250, the city's walls were demolished, and were subsequently rebuilt city from the Ottomans during World War I and it was to be included in an international zone under the 1947 United Nations Partition Plan for Palestine occupied by Israel in the 1967 Six-Day War. Since 1995, Bethlehem has been governed by the Palestinian National Authority. Bethlehem has a M communities. The Bethlehem agglomeration includes the towns of Beit Jala and Beit Sahur, as well as the refugee camps of Aida and Azra. Bet Christmas season when Christian pilgrims throng to the Church of the Nativity. Bethlehem has over thirty hotels and three hundred handicraft work northern entrance of Bethlehem.Belén "Ciudad Natal de David" (árabe بيت لحم "casa de la carne"; hebreo בית לחם) es una ciudad palestina situada al sur de Jerusalén, enclavada en actualidad por la Autoridad Palestina. Su población es de 27 000 habitantes, siendo la mitad musulmanes y la otra mitad cristianos, en su mayor musulmanes al ser, de acuerdo con la Biblia, el lugar de nacimiento de Jesús de Nazaret según los evangelios de Lucas y Mateo. Es también un Raquel situada a la entrada de la ciudad y para los que la ciudad es lugar de nacimiento y de coronación del rey David. El gobierno de Israel ha ro terroristas, impidiendo el libre tránsito de los habitantes y limitando los intercambios comerciales. Esto ha provocado una gran disminución del tu Bethlehem tai Bethlehem (arabaisi: بيت لحم, bayt lahm) on 20 000 asukkaan (1993) kaupunki Palestinan Länsirannalla. Bethlehem sijaitsee kahdeks pyhimäeläiskohde, jossa järjestettävään joukkumääränpalvelukseen osallistuu vuosittain suuri joukko paikallista kirstitystä ja pyhimäeltä. Bethlehem vuoksi: Huhtikuussa 2002 Israelin armeija miehitti Länsirannan kaupunkeja, mt. Bethlehemin, operaatio Defense Shieldin yhteydessä. Talon Jeru palestinianantautiepile tuntepakon Jeeuskaan syntymäkohissaBethlemme è una città della Cisgiordania, capitale del Governatorato di Bethlemme dell'Autorità Nazionale Palestinese, ed è famosa soprattutto per Gesù Cristo: per questo motivo vi sorge la Basilica della Natività. Si trova a circa 10 km a sud di Gerusalemme, ad un'altezza di 765 m sul livello c Jala e Beit Sahur. Secondo la Bibbia, a Bethlemme nacque Davide, secondo re di Israele; essa è quindi menzionata da Luca evangelista c essere suo discendente e nascere nella sua città. Secondo i Vangeli, la nascita di Gesù adempi questa profeta. Bethlemme (بيت لحم) significa "C petribe anche significa "Casa della carne" se viene tradotta dall'arabo (بيت = "Casa"), ma non mancano altre etimologie che fanno riferimento stella di Bethlemme viene comunemente chiamata stella cometa.ベツレヘム (Template:Lang-ar (Bayt Lahm) バイト・ラハム、ベト・レヘム、[拜]の 前に位置する都市である。日本では正教宗において「ベツレヘム」とも表記され、Bethlehemon? بيت لحم in Palestinese Autoriteit Bestand Flag of Pal Algemeen Inwoners (2006) 29 330 Hoogte 765 m Portaal Bestand Postal code aangrenzende Beit Jala en Beit Sahur hebben respectievelijk 13 000 en 12,1 belangrijke mate afhankelijk zijn van hulp van de VN. Bethlehem ligt ongeveerBethlem eller Bethlehem er en palestinsk by på den okkuperte Vestbredden for Jerusalem. Byen er viktig innen kristendommen siden den felge Bibelen eBethlem - miasto w środkowej części Palestyny. Położone w Judei, na obszar

→ <http://dbpedia.org/resource/Bethlehem>
(Bethlehem URI in DBPedia)

GeoNames Home | Postal Codes | Download / Webservice | About

Map center: N 31° 42' 48" E 35° 12' 1"



displaying Geonames id : 284315
search for display all features in area

GeoNames Wiki/edia

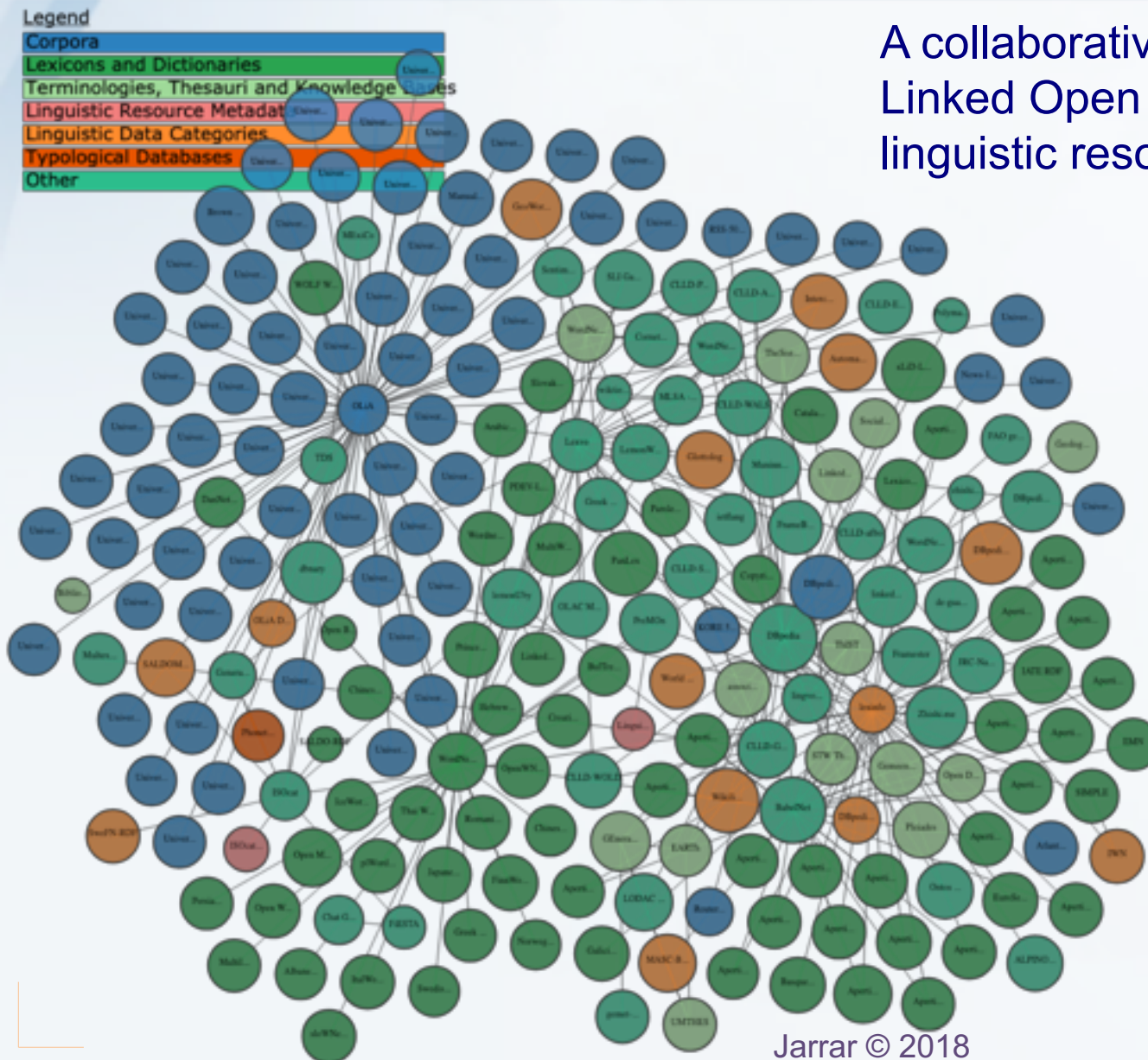


<http://sws.geonames.org/284315/> →
(Bethlehem URI in Geonames)

Linguistic Linked Open Data Cloud

A collaborative effort to develop a Linked Open Data (sub-)cloud of linguistic resources.

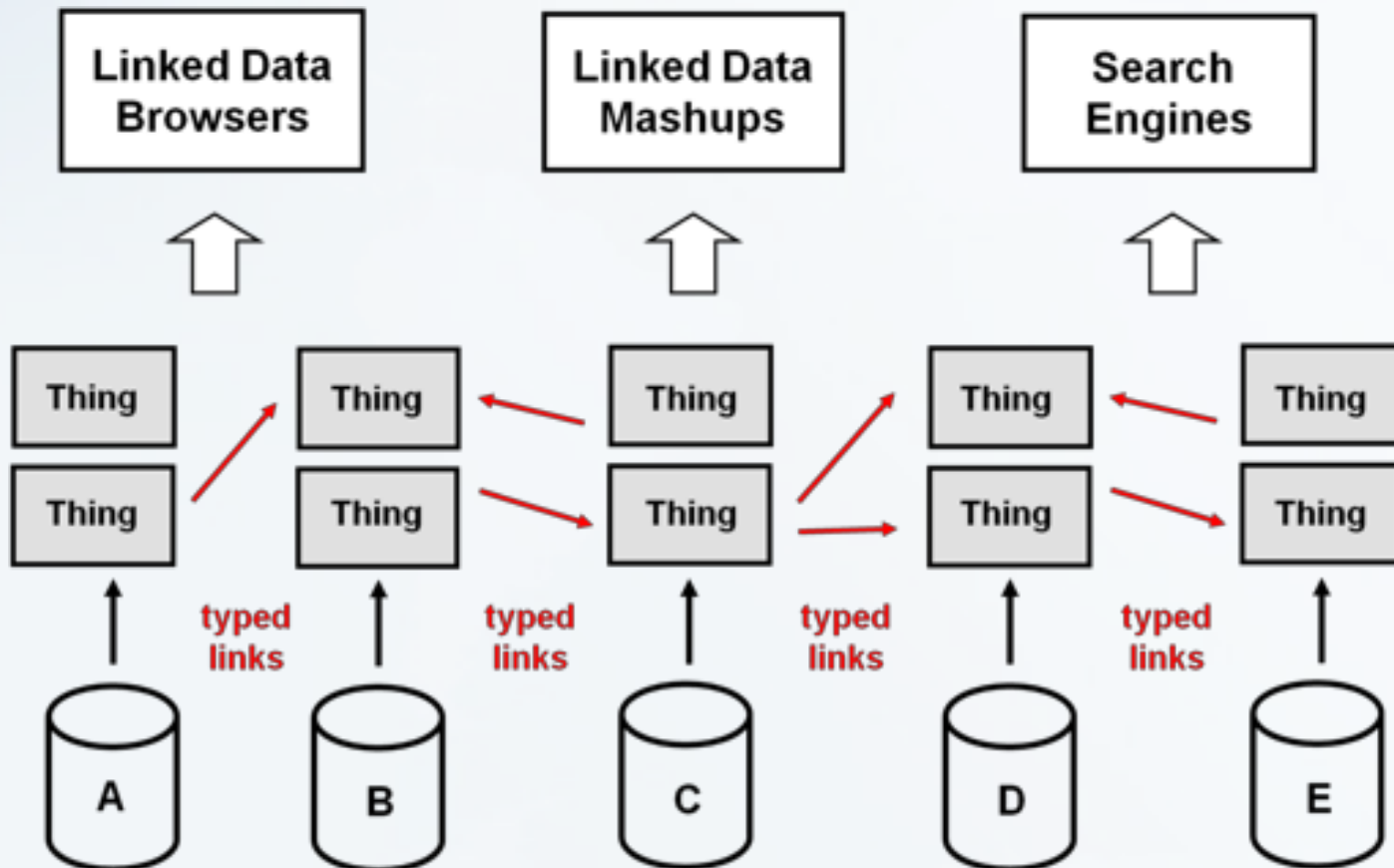
Represent (lexical entries, concepts, synsets, and other) using Lemon RDF model, then interlinked.



Applications

Based on [1]

What can I do with this?



Ideas for projects

Each student can (1) download data about entities (e.g., related to Ramallah) from the following sources (but you may find other sources), (2) represent it in RDF, then (3) link his/her data with other student(s):

Google Maps
Google street maps
GeoNames

Google Scholar
Research Gate
ORCID
PubMed

WikiData
YouTube
Twitter

Some students can build a store to publish data from all students allows people to search/query it.

References

1. Bizer, Christian, Tom Heath, and Tim Berners-Lee. "Linked Data - The Story So Far." *International Journal on Semantic Web and Information Systems*: 1-22. Print.
2. Christian Bizer: Lecture notes on The Emerging Web of Linked Data
<http://www.dima.tu-berlin.de/fileadmin/fg131/FOKO/Bizer-TUBerlin-LinkedData-Talk.pdf>
3. Mustafa Jarrar, Anton Deik: **The Graph Signature: A Scalable Query Optimization Index for RDF Graph Databases Using Bisimulation and Trace Equivalence Summarization.** *International Journal on Semantic Web and Information Systems*, 11(2), 36-65, DOI: 10.4018/IJSWIS.2015040102. April-June 2015
4. Anton Deik, Bilal Faraj, Ala Hawash, Mustafa Jarrar: **Towards Query Optimization for the Data Web - Two Disk-Based algorithms: Trace Equivalence and Bisimilarity.** Proceedings of the 3rd Palestinian International Conference on Computer and Information Technology (PICCIT 2010). Hebron, Palestine. March 2010.
5. Mustafa Jarrar and Marios D. Dikaiakos: **A Query Formulation Language for the Data Web.** *IEEE Transactions on Knowledge and Data Engineering*. IEEE Computer Society.
6. Mustafa Jarrar, Marios D. Dikaiakos: **Querying the Data Web: the MashQL Approach.** *IEEE Internet Computing*. Volume 14, No. 3. Pages (58-670). IEEE Computer Society, ISSN 1089-7801. May 2010. Mustafa Jarrar and Marios D. Dikaiakos: **A Data Mashup Language for the Data Web.** Proceedings of LDOW, WWW'09. ACM. ISSN 1613-0073. (2009).
7. Mustafa Jarrar and Marios D. Dikaiakos: **MashQL: a query-by-diagram topping SPARQL -Towards Semantic Data Mashups.** Proceedings of ONISW'08, part of the ACM CiKM conference. ACM. pages (89-96) ISBN 9781605582559.(2008).